

## ***ClinkNotes*: Towards a Corpus-Based, Machine-Aided Programme of Translation Teaching**

Chunshen Zhu et Po-Ching Yip

Volume 55, numéro 2, juin 2010

URI : <https://id.erudit.org/iderudit/044247ar>

DOI : <https://doi.org/10.7202/044247ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)

1492-1421 (numérique)

[Découvrir la revue](#)

Citer cet article

Zhu, C. & Yip, P.-C. (2010). *ClinkNotes*: Towards a Corpus-Based, Machine-Aided Programme of Translation Teaching. *Meta*, 55(2), 387–408.  
<https://doi.org/10.7202/044247ar>

### Résumé de l'article

Le présent article fait l'état des lieux d'un projet pilote relatif à la création d'une plateforme conçue pour l'enseignement de la traduction ou la formation bilingue, à grande échelle, aux études supérieures. Bien que les premiers textes utilisés dans le cadre du projet soient en anglais et en chinois, le programme, *ClinkNotes*, offre la possibilité de prendre en charge des corpus parallèles de n'importe quelle paire de langues. L'article débute par un bref survol de l'application des corpus à la traductologie en lien avec la formation professionnelle en traduction. Puis les caractéristiques du programme (cadre théorique, méthode d'annotation et fonctionnement) sont présentées, ainsi que la manière dont il comble les impératifs pressants de la profession. Les perspectives futures d'amélioration du programme sont également discutées.

# ***ClinkNotes*: Towards a Corpus-Based, Machine-Aided Programme of Translation Teaching**

**CHUNSHEN ZHU**

*City University of Hong Kong, Hong Kong, China*  
ctzhu@cityu.edu.hk

**PO-CHING YIP**

*City University of Hong Kong, Hong Kong, China*  
dryippoching@yahoo.co.uk

## **RÉSUMÉ**

Le présent article fait l'état des lieux d'un projet pilote relatif à la création d'une plateforme conçue pour l'enseignement de la traduction ou la formation bilingue, à grande échelle, aux études supérieures. Bien que les premiers textes utilisés dans le cadre du projet soient en anglais et en chinois, le programme, *ClinkNotes*, offre la possibilité de prendre en charge des corpus parallèles de n'importe quelle paire de langues. L'article débute par un bref survol de l'application des corpus à la traductologie en lien avec la formation professionnelle en traduction. Puis les caractéristiques du programme (cadre théorique, méthode d'annotation et fonctionnement) sont présentées, ainsi que la manière dont il comble les impératifs pressants de la profession. Les perspectives futures d'amélioration du programme sont également discutées.

## **ABSTRACT**

This article presents a report on a pilot project designed to construct a platform for large-scale teaching of translation or bilingual training at tertiary level. The programme, *ClinkNotes*, has the potential of accommodating parallel corpora of any language pairs, although the primary data used in this project are in English and Chinese. The report begins with a brief overview of the development of corpus-based approach to translation studies in relation to that of translation teaching as a profession. It then proceeds to describe the actual design (i.e., the theoretical framework, the methodology of annotation, and the simple execution of the software programme), and how it helps to cater to the pressing needs of the profession. The prospects of further development of the programme are also discussed.

## **MOTS-CLÉS/KEYWORDS**

pédagogie de la traduction, corpus parallèles, plateforme d'enseignement  
pedagogy of translation, parallel corpora, teaching platform

## **1. Introduction**

### ***1.1. Translation teaching and the need to increase pedagogical efficiency***

As observed by Brisset (2003: 103), since the end of World War II and the establishment of international organizations such as the United Nations, which demand large quantities of translation on a daily basis, the world has seen a steady process of professionalization and industrialization of translation. In response to such market demands, the past several decades have witnessed the rise of an academic discipline called Translation Studies and an increasing number of translation courses and

programmes offered at universities across Europe, Canada, Australia, and China, among others. In China, for instance, almost every (foreign) language programme in its thousands of universities has a translation component, and the government is reported to be piloting an undergraduate degree curriculum in translation in some of the country's leading foreign language institutes.<sup>1</sup> In Hong Kong, a former British colony and now a Special Administrative Region of the PRC, the degree programme in translation between English and Chinese, at both under- and post-graduate levels, has for nearly half a century been among the most sought-after programmes in its universities.

Yet, due to its nature of multiple acquisition of languages, knowledge and skills, the teaching of translation has always remained a labour-intensive, time-consuming, and space-constrained pedagogical endeavour. Let us take the City University of Hong Kong, where the first author is affiliated, for example. According to the department's records, the annual intake of its three-year government-subsidized programme of B.A. in translation and interpretation is 53 in 2005, while that of its Translation/ Interpretation specialization on its newly launched self-financing programme of B.A. in Language Studies is 80. That means each year on campus there can be as many as 399 full-time students working on a translation (or translation-related) B.A. degree programme, apart from those taking translation courses on a part-time M.A. programme in Language Studies. The number may predictably increase by one third by 2012, when a normative four-year curriculum is introduced in Hong Kong's tertiary education.

Teaching of translation on such a large scale needs rigorous mechanisms for quality assurance in the first place. How can the tutor responsible for teaching the class manage to engage those several hundred motivated individuals all at once, satisfying their variously-disposed inquisitive minds, and providing an adequate and respectable answer to every question of theirs? Teaching translation under such circumstances can become an enterprise that is intrinsically unsympathetic to individual needs and susceptible to prescriptive generalities (however unintentional). Inexperienced teachers may even unwittingly fall victim to their own naively-empirical comments, explanations, or solutions offered offhandedly on the spur of the moment. More frustrating still are cases where a tutor was caught answering the same question over and over again obviously in want of theory-informed consistency when it was put to him by different students at different levels on different occasions. This being the case, is there a viable means to remedy the current unwieldy situation or at least to supplement, if not to supplant, the restrictive (though appreciably humanistic) classroom tradition, for example, to open a channel that not only encourages self-tuition but is also pedagogically efficient and cost-effective in the long run?

Moreover, to optimize the efficacy of the training, a student-centred approach in teaching, or student autonomy in learning, seems to be the order of the day. And there need to be measures in place to enable trainees to "transcend the limitations and constraints of their immediate learning environment" (Little 2000: 70), and knowledge to be "constructed by learners, rather than being simply transmitted to them by their teachers" (Kiraly 2000: 1).

### 1.2. *Machine, translation, and the teaching of translation*

If the professional practice of translation today is to be marked by an emphasis on theory-backed “quality assurance” as advocated by Brisset (2003: 103), then, in the large-scale teaching of translation and the training of professional translators, a corresponding emphasis on efficiency-motivated, theory-informed “quality assurance” should also be introduced.

To address these concerns of quality and efficiency in translator training, can machines help? This is our question.

Machines have indeed helped, so far as translating itself is concerned and mainly on two fronts: one is Machine Translation (MT), an undertaking, which might have set out with an ambition to relieve humans completely of the burden of translation, has now turned from MT to a more realistic mode of CAT – computer-aided translation that involves post-editing by the human translator – a convenient combination of artificial and human intelligence, so to speak. And the other is the corpus-based approach to translation studies, apparently a cause celebre in the discipline made possible by the sophisticated technological advance in computers and related software.

The sudden interest in this newly-found research tool is not without rhyme or reason. First of all, computers have made it possible for vast quantities of text to be readily comparable, manageable, storable, and extractable at the touch of a button whenever necessary and for whatever designated purposes. And then Baker (1993), a translation scholar and theorist, was quick to latch on to this corpus-linguistics-informed, computer-technology-enabled development in translation studies in her seminal paper, where she predicts enthusiastically: “[t]he profound effect that corpora will have on translation studies, in my view, will be a consequence of their enabling us to identify features of translated text which will help us understand what translation is and how it works” (Baker 1993: 242-243).

Since then, the corpus approach has been applied in translation studies as a “viable and fruitful perspective” and “a novel and systematic way” of research which “addresses a variety of issues pertaining to theory, description, and the practice of translation” (Laviosa 1998: 474).

However, the corpus approach has so far focused more on the identification of “features of translated text” (Baker 1993: 243) in an attempt to systematically theorize and describe (the practice of) translation *per se*. Whilst performing certain humanly impossible tasks to generate new insights, its scope is somewhat constrained by its mode and level of data mark-up, tagging, processing, and retrieval which tends to follow a lexico-grammatical paradigm, as is often seen in the practised manoeuvre of word/phrase alignment, part-of-speech tagging, word counting, and keyword-in-context concordancing, to indicate such linguistic tendencies as frequency of occurrence of single types, type-token ratio, and lexical density (Kenny 1998).

It is true that corpus approaches at present are powerful in tracing out “what translation is and how it works” (Baker 1993: 243) in such aspects as translational behaviour, equivalence relationships, and translationese with “[t]ypical applications [including] translator training, bilingual lexicography and machine translation” (Kenny 1998: 51). However, in frontline teaching of translation, we are still in need of a methodology that will enable us to apply machine-aided and corpus-based

approaches in a systematic and rigorous fashion to the improvement of the *pedagogical efficiency* by addressing such issues as “the need to explain the reasons for the advice given [to learners]” (Lederer 2007: 22), the cultivation of a learner’s awareness of cultural sensitivity, the importance of background and professional knowledge in translation, the correlation between the text and its intended effect, the diversity of the methods used in producing a target text, and so on. In our opinion, the explanation of an act of translation, as well as the act itself, is of equal importance and should also demonstrate a degree of consistency informed by theory and research, and should not be a hit-and-miss exercise dictated by the translator’s intuition alone. Therefore, to complement existing approaches and bring corpus research to the aid of translation pedagogy, it is time we pursued a new line of investigation as Bowker suggests in her discussion of a corpus-based approach to translation evaluation, “more sophisticated extraction techniques” have to “be developed for use on corpora that have been part-of-speech tagged or annotated in other ways” (Bowker 2001: 361).

The programme *ClinkNotes* reported in this article is precisely designed to cater for such pedagogical needs. As the name suggests, it is basically a programme that enables the user to “click to link to the notes,” which are grouped and prepared under two separate systems: one on cultural and background knowledge and the other on methods used in the actual translation of a text. The features of the programme are described in the following sections.

## **2. *ClinkNotes*: a platform for efficient teaching of translation**

### **2.1. *Modes of annotation and retrieval***

#### **2.1.1. *Click to link up with notes: a dual mode of annotation***

As the difficulty of translation mainly comes from two aspects, i.e.:

- 1) access to the cultural and intertextual background of a literary source text, or for that matter to the subject and professional knowledge needed for understanding a text of a more pragmatic nature, e.g., a financial or medical text;
- 2) the textual accountability between the source and target texts in terms of function and effect;

the corpus is annotated in a dual mode to deal with problems from both aspects, which is described below.

#### **2.1.2. *Annotating the corpus by the paragraph***

As shown in Table 1 in Appendix, the source and target texts are aligned side by side by paragraphs in table form. Each paragraph of the source text and its corresponding target text (normally a matching paragraph) are numbered at the beginning according to its place in the paragraph sequence (e.g., the number [23] in Table 1 indicates the 23<sup>rd</sup> paragraph of the text), forming an independent unit for annotation purposes. That is, all the annotations in a particular paragraph are marked with a common prefix, which is the number of the paragraph, for example [23], for all annotations in Paragraph 23. This helps to keep to a minimum the knock-on effect of adding or deleting an annotation on the serial numbering system either in the annotating process or during the trial run.

### 2.1.3. Annotations on background knowledge

As the corpus used for this programme is culturally rich rather than science- or technology-oriented, those details in the source text that are deemed to require further information on their cultural or intertextual background for ease of cross-cultural comprehension are first identified, highlighted in bold, and then annotated. A knowledge annotation of this type is marked, in the source text only, at the end of the annotated sentence(s) in a format of [*paragraph number . serial number*]. For example, the [24.1] in Table 1 indicates the first knowledge annotation in Paragraph 24. By the same token, the second knowledge annotation in Paragraph 7 is marked as [7.2].

### 2.1.4. Annotations on translation methodology

Similar to knowledge annotations, those notable details in the target text, in the light of their counterparts in the source text, that illustrate one or more translation methods are first identified, boldfaced in the English text and underlined in the Chinese, and then annotated. However, a methodology annotation of this type is marked at the end of the annotated sentence(s) in both target and source texts, and in a different format of [*paragraph number followed by a letter from the English alphabet*]. For example, the [23b] in Table 1 in Appendix indicates the second methodology annotation in Paragraph 23 in both texts. By the same token, the first methodology annotation in Paragraph 24 of the texts is marked as [24a].

It has to be noted that the selection of details for annotation can be a rather arbitrary practice, but in annotating our present corpus of Oscar Wilde's *De Profundis* and its Chinese translation, thanks to annotated editions of the source text such as Hart-Davis (1962) and Hyde (1982), the project has a reasonably balanced collection of sources of information and references for data verification at various stages of annotation. The device of numbering annotations on a paragraph basis has rendered the system as a whole amenable to being edited, revised, modified and upgraded by adding new annotations with little knock-on effect or without the need of further programming.

It is understandable that a teacher may welcome as many annotations as deemed necessary to facilitate his/her teaching, but in actual practice, heavy presence of annotations, particularly in the form of footnotes or endnotes in printed books, can sometimes become too much of a distraction, putting a positive reading experience in jeopardy. In this connection, the machine by comparison is far more helpful in being able to first hide the notes and then allow the user to retrieve them by clicking on any annotation marker or searchword as required. And more significantly, as we shall see, a machine-aided programme as such can group together annotations pertaining to the same translation phenomenon for the purposes of more focused deliberation and research.

Also, to encourage a more proactive way of using the programme, for each translation method or phenomenon identified, only selected examples are marked and annotated with a view to illustrating the method or phenomenon concerned rather than to exhausting its manifestations throughout the corpus. The selection of examples may therefore range from as few as 1 (e.g., for *Capitalization*) to as many as, say, 58 (e.g., for *Word Class Shift*), depending on the usefulness and frequency of

the methods or phenomena thus discerned. In other words, through such annotations, learners are encouraged to identify further interesting examples for themselves within and beyond the corpus, after being alerted to the method or phenomenon in question. Moreover, the annotations are there simply to prompt learners to come up with their own translations and explain them with reference to the existing translation and its analytic justifications by way of annotation.

### 2.1.5. *Multiple modes of access to annotated examples*

The system can be installed as a document folder in *Microsoft Word*, with the *Macros Security* set at *medium*. Users click into the folder and click on the file *cover* and then *Enable Macros* to start the operation. A cover page will then appear on the screen. Basically, once the text page is clicked open, all knowledge and methodology annotations can be easily retrieved by right-clicking on their respective annotation markers as illustrated above and in Table 1 in Appendix. A user's manual is also available when the *Instructions* box at the bottom of the cover page is clicked on. It takes the user through all the accessible functions associated with the corpus and explains every operation in simple and straightforward terms.

### 2.1.6. *Instant access to notes via the searchword list*

As noted above, annotations of both knowledge and methodology types can be accessed by clicking on the appropriate marker in square brackets as the user goes through the texts and finds a translation phenomenon interesting. Besides this, the corpus always opens with a smaller window containing a list of searchwords indicating relevant translation methodologies or phenomena popping up in the centre, which can be moved around the screen (see Figure 1 in Appendix), and can be hidden and recalled at the user's will by a simple click of the mouse. The user, once in the corpus, can retrieve example(s) under a particular searchword directly from the list by clicking on the searchword, e.g., *allusion*. Examples retrieved are accessible *via* an annotation window (see Figure 1). While Figure 2 in Appendix is a background knowledge annotation window shown against the background of the bilingual corpus, Figure 3 in Appendix is a methodology annotation window with details about its operation.

### 2.1.7. *Quick move from one note to another*

As various steps in Figure 3 indicate, with such information as the number and locations of examples provided, the user can retrieve any of the examples annotating a particular translation phenomenon by clicking either on the searchword in the drop-down list (step 2) or on a round-bracketed searchword embedded in the annotation itself (step 1).

### 2.1.8. *Grouping of examples and annotations*

For various purposes, the user may want to group and show all the examples under a particular searchword. This is made possible by clicking on the *Show All* button in the annotation window. The examples will be shown with their markers and annotations in a new *MsWord* document (see step 3 in Figure 3 and Figure 4 in Appendix).



### 2.1.9. *Instant reference to the context*

This function of grouping examples in a new document can be particularly useful in classroom teaching. While being a separate file, however, it is still connected to the main programme, so long as the latter remains on. The connection enables the user:

- 1) to instantly refer a particular case back to its context, for example, click on its methodology marker (e.g., [3g]) and the system will automatically show the bilingual passage highlighted in the corpus;
- 2) to retrieve a background knowledge annotation by clicking the appropriate marker as if it were in the main corpus (see Step 1 in Figure 4, and then Figure 5, in Appendix).

### 2.1.10. *Screen management for classroom teaching purposes*

To facilitate classroom presentation for instance, the user can click on the asterisk button on the window to maximize its size or reduce it to the original. Also, an X button can be clicked to close a window.

A comprehensive manual is always available as an integral component of the system for users' ease of reference with a click of the *Instruction* button on the cover, or, when in corpus, with a double click on the left button of the mouse.

## 2.2. *Selection of the parallel corpora*

The system outlined above is designed to serve as an initial platform on which corpora of different genres and types of subject matter can be annotated and electronically managed for teaching and learning purposes. According to our long-term plan, it will develop into an open series of independent programmes, covering more than one translator and language pair, with each of the programmes accommodating a particular type of text or subject matter, such as science and technology, finance, news, and fiction. For the present project, which, being the pilot, is meant to explore the ground and gauge the degree of delicacy of annotations for the system to be viable, we chose the full text of Oscar Wilde's *De Profundis* with the Chinese translation by the first author. This was done for the following reasons, apart from copyright considerations.

1. Being a text of around 50,000 words, it is comprehensive yet manageable in helping alert users to the textual cohesion and coherence as reflected through translation, e.g., cross-paragraph repetition, semantic echoing, and co-reference. This can be regarded as an advantage over a corpus consisting of a host of discrete passages from different sources.
2. Being a culturally charged text, it represents a text that is sufficiently complex in textuality and intertextuality, extensive in its topical coverage, rich in its stylistic presentation, and hence, as we shall see, challenging for a translator. In other words, the corpus can be hoped to illustrate the most advanced level of difficulty in terms of language, rhetoric, style, cultural intertextuality, and translation methodology.

Of course one may contend that, being a private letter written in 1897, the text is irrelevant to the world of our present-day students, or that being a stylist, Oscar Wilde was writing in a language that is too *literary* to be pertinent to the language reality of a professional translator.



Such concerns can be explained away as follows. While the detachment of the subject matter from the present-day world may serve to highlight textual issues involved in translation, the choice of an author like Oscar Wilde into the bargain was in fact a plus. For being an avowed stylist, he exploits the mechanism of the English language to its very limits, which could pose a formidable challenge to a translator, especially when the translation is to be a work of comparable quality annotated alongside the original in a parallel corpus. Besides, *De Profundis* (Wilde 1897/1951), an exceptionally long letter which Wilde wrote in prison with its 173 paragraphs representing an immensely rich reservoir of stylistic and topical diversities, is adequate enough to build into a sizeable yet coherent parallel corpus for our purposes at hand.

Admittedly, the source text sounds *literary*, especially when it broaches an artistic or literary topic, or when the author vents his feelings in a pungent and, if you like, *poetic* way. But, as argued extensively in the literature, literary texts and pragmatic texts, though they may differ along a cline or gradation of *literariness* (Carter and Nash 1983: 124) or in their linguistic artistry (Leech and Short 1981: 2, 6), do share a high degree of textual similarity in their use of language; or as Cook (1994/1999: 2) points out: “Many literary works, on the contrary, seem pointedly to borrow the language of non-literary discourses.” It therefore follows that the study of literary translation and that of non-literary translation are interrelated (see also Zhu 2004, especially section 3 for a discussion of the relevance between literary and non-literary translations). But we have not lost sight of the fact that in literary works there is “something accessible, beautiful, understandable, enjoyable, and uplifting” contributing to “a common experience which cuts across the boundaries of nation, culture, and history” (Cook 1994/1999: 3).

When considered from the perspective of translation methodology, we feel it is not unreasonable to surmise that a literary text

- 1) encompasses a wider range of linguistic, cultural, social, political, historical, religious, and stylistic variations and references waiting to be ingeniously tackled in translation;
- 2) constitutes greater challenges for the translator linguistically and otherwise;
- 3) explores a much richer (and perhaps more inspired and inspiring) variety of methodological strategies in writing and in translation alike.

If so, we may safely assume that by researching the translation of a style-rich original we shall be able to identify and retrieve a superordinate array of translation strategies, the greater part of which, due to the textual similarity of literary and non-literary texts, should in theory be found to be applicable to, and applied in, the translation of texts of other types by a host of other translators. Incidentally, the list of searchwords has been used in a similar programme (Zhu 2005) which incorporates 30 English-Chinese pairs of texts from the *Financial Times* website as the basic corpus; and the list, while remaining open, is actually designed to be the unifying component in the follow-up projects of the planned series.

For the Chinese target text, we use the first author’s translation 自深深处 (*zi shen shen chu* literally *from a deep deep place*) simply because this text is the latest translation of *De Profundis* and it has been used by the translator himself in his teaching.<sup>2</sup>

We understand perfectly that despite its provenance and the convenience regarding copyright clearance, the use of one’s own work may pose a serious ethical issue,

namely, the possibility of biased subjectivity and ungrounded self-justification. To obviate such situations, controlling mechanisms have to be installed in our *modus operandi* to guarantee the objectivity of the investigation and the accountability of the annotation.

One of such mechanisms is the formulation of a coherent theoretical framework informed by discourse linguistics and stylistics, which is actually the most outstanding feature of this project. The other is the crucial involvement of the second author as an informed and independent critic in the annotation of the translation methodology.

Nonetheless, it has to be stressed that no translation is supposed to be final or perfect nor is the project designed to cover every single translation technique used. In actual use, the corpus is always presented as an object for critical deliberation and a source of inspiration for better informed choice and more sensible adoption of translation methodologies.

### 2.3. *Categorization of methodological phenomena via searchwords*

As a machine manageable pair of texts, the annotated corpus is hoped to provide learners with illustrations of translation strategies retrievable in an organized fashion. One of its strengths, therefore, lies, not in the size of the data it incorporates (which, incidentally, is relatively small compared with mainstream databases that would be capable of embracing collections of millions of words), but in its theory-informed observation and systematic categorization of translation methodologies: a list of 225 items has been compiled in the light of information structuring, and in its application of such observations to the actual teaching of translation in an electronically interconnected way.

The list of searchwords is of course an open-ended compilation, augmentable to cover newly found methods or phenomena as the corpora expand. In the eventual published edition, it will appear in the printed book as a glossary, which outlines a workable scheme of categorization in relation to methodologies or cross-language phenomena based on focus-sensitive information structuring. This conceptual framework, which forms the overarching backbone of the system's annotations, is explained in the subsections below.

#### 2.3.1. *An information-oriented theoretical framework*

As we pointed out above, the present project proposes an approach complementary to the mainstream lexico-grammatical paradigm in corpus-based research by focusing on textual management across languages, especially in terms of the relationship between information structuring and discourse function as seen in the context of translation. As Lambrecht (1994: 338) observes, "Discourse function is [...] inherent in the formal system. Sentences do not exist without information structure [...]." In our attempt to account for the translation phenomena in the annotation, the informational approach of theme-rheme in systemic-functional linguistics (Halliday 1994), coupled with Lambrecht's topic-focus approach has provided us with a feasible point of departure.

In a nutshell, the working hypothesis derived from the literature is as follows. According to Halliday, in an *unmarked* information structure of theme-rheme

sequence, the theme, which is in the clause-initial position, represents given information, while the rheme encapsulates the new information that is to be conveyed in the communication. In Lambrecht's model of information structure, *theme* plays a *semantic role*, and is to be viewed as *topic* when it designates a *pragmatic relation* to *focus* (Lambrecht 1994: 15, 342 Note 11). Taken together, theme becomes the pragmatic topic, and the focus, which occupies the clause-final position (known as the *end-focus*), becomes pragmatically the most accentuated part of the rheme. However, apart from end-focusing, an ideational entity can be marked as a focal point through other accentuating devices such as the passive voice and cleft constructions, and modification (Zhu 1996a, 1996b). In other words, either topicalizing or focalizing an entity underscores its informational importance and, as such, creates a communicative impact on the receiver, which should have a pragmatic bearing on a translated text. The methodology annotation, in essence, is an attempt to account for a translator's awareness of various manifestations of *markedness* in both source and target texts.

Such a hypothesis has elevated our investigation from the level of form, i.e., a mere observation of the transfer of a language's formal characteristics, to that of function in terms of the encoding and decoding of information through the use of language. Thus when comparing the original and the translated texts sentence by sentence, we are able to account as much as possible for the differences and similarities (apparent or minute) in information structuring where the source and target languages are seen to be relating the same situation or event. In this way, we may perhaps succeed in registering the multivarious yet systematic workings of the human mind in the very process (conscious or subconscious) of information transaction through writing, and likewise through translating. In other words, all the linguistic manoeuvres (syntactic, lexical, or stylistic) carried out by the writer in the source language and matched by the translator using different or similar linguistic tactics in the target language are viewed through this self-same theme-rheme prism as tangible expressions of the same intended information being processed in language-specific ways.

It has to be noted, however, that this programme is not meant to be an exercise to spell out linguistic complexities in detail. Instead, it incorporates in its framework for methodology annotations such basic notions of information structuring as characterized in discourse linguistics. To follow the annotations, users are not expected to be *linguistic experts*, so long as they can familiarize themselves with these fundamental notions, or *theoretical principles* in Lederer's (2007) terms, as are explained in the glossary provided.

### 2.3.2. Searchwords and searchword-oriented translation methods and phenomena

Different manifestations of *markedness*, or different linguistic manoeuvres, discernible in our data are tagged or *oriented* by searchwords as identifiable translation methods, or in more general terms, as information-related translation phenomena. These methods are conceptually categorized as follows with reference to their most outstanding features in terms of information structuring:

- 1) Information organization;
- 2) Information distribution;
- 3) Information realization;
- 4) Information representation;
- 5) Information explicitation;
- 6) Information and para-information;
- 7) Information reformulation.

These categories are characterized and illustrated in the following subsections.<sup>3</sup>

### 2.3.2.1. *Information organization*

In a broad sense, every textual phenomenon can come under this heading of *information organization*. Here, however, we are using the term to refer only to textual phenomena in terms of theme-rheme or topic-focus sequence. Such a perspective on the organization of information is what McCarthy terms *framework*:

In English, what we decide to bring to the *front* of the clause (by whatever means) is a signal of what is to be understood as the *framework* within which what we want to say is to be understood. The rest of the clause can then be seen as transmitting 'what we want to say within this framework' (McCarthy 1991: 52, original italics).

We shall see to what extent this concern with *framework* and its textual effect can prompt a Chinese translator to tap his target language for linguistic devices to create a desirable impact. This can be illustrated by the following two different translations of the same cleft structure, the markedness of which, according to Kenny (2001), can either be nullified or retained in translation.

- (1) It is man's soul that Christ is always looking for<sup>[100a]</sup>

which could have been translated conveniently into an unmarked or normalized structure as 基督一直在找寻人的灵魂 (literally *Christ always is looking for man's soul*). Yet to reproduce the effect of markedness, the translation provided in the corpus reads 基督一直在找寻的是人的灵魂 (literally *Christ always is looking for de shi man's soul*), which employs an equally marked *de shi* structure in Chinese (Zhu 1996a), with the *fronted focus* being changed to *end weight* without losing the intended *emphasis*. This is explained in the annotation with further links to other searchwords indicated in bold (reproduced here in italics).

To accentuate the same marked entity as is encoded in the source text seems to be particularly essential in cases where retaining the same focal point is deemed relevant and important to the coherence of the text. As a matter of fact, the formal sequencing of theme and rheme using varying grammatical constructions seems to be usually dependent on the coherence requirement of the context in which the sequence occurs. Any inadvertent change to the arrangement in the translation will inevitably affect the ordered flow of information in the process, and will therefore interfere with a hearer-reader's assumptions and expectations (such interference, to be sure, may be said to be intentional in certain circumstances).

Methodologies (in italics) covered under this category are: [the entity of theme/rheme] *unchanged* or *reversed*; partial theme-rheme *combined*; *fronted focus*, including *it is construction* and *what construction* in English, *shi construction* in Chinese, and other contextually-induced devices of *emphasis* in both; *end weight* or *end focus*;

unchanged or reversed *clausal sequence*; *coherence* in relation to *perspectives*, *syntactic cohesion*, *lexical cohesion*, or *prosody*.

### 2.3.2.2. Information distribution

As information generally unfolds *via* language in a linear fashion, there are always optimal ways of assigning different parts of the information to various focal or non-focal points in a linguistic formulation. Such distributional operations, however, are often dictated by stylistic decisions, lexico-syntactic obligations, or focal necessities; and to some extent constrained by the existing repository of a language's vocabulary, word classes and syntactic compatibilities. This is especially the case in translation as conventional and predilectional ways of arranging information are not often identical between languages. Sometimes a sequence-marked focal entity, for example, will have to be recast in a lexically-marked way to overcome the constraint of linearity, as it were. In the following example,

- (2) [o]f seed-time or harvest, of the reapers bending over the corn, or the grape-gatherers threading through the vines, of the grass in the orchard made white with broken blossoms, or strewn with fallen fruit, we know nothing, and can know nothing<sup>[49c]</sup>.

春种秋收，农人在田里俯身挥镰，果农穿行于藤蔓间采摘葡萄，果园的青草上，残花落时一片片的白，果子掉下又散散的滚了一地：这一切，我们一点也不知道，一点也无法知道<sup>[49c]</sup>。

(lit.: *spring sowing autumn harvest, farmers in the fields bend and wield sickles, fruit-growers thread through vines pick grapes, on the orchards' green grass, when withered flowers fall patches of white, [when] fruit drop [they are] strewn all over the ground: all this, we not at all know, no way at all know*)

where, as annotated, the phrase *all this* has been added, which not only serves to summarize the preceding information, but also helps to front the object so as to conform to the marked (or topicalized) information introduced by the fronted *of*-phrases in the source text – a linguistic operation hardly possible in the idioms of Chinese.

The following searchwords (in italics) indicate tendencies of this kind: *information flow* similar to the source text; *analytic redistribution* without missing out any details; *multiple redistribution* through multiple lexical units in a complementary fashion, e.g., *idiom for word*; *condensation* through simplification, e.g., for the sake of *brevity*; *collocation*; and *summary recapitulation*.

### 2.3.2.3. Information realization

Information realization is concerned with language-specific devices available in, or preferable by, a language to register a piece of information. In translation, these linguistic forms or conventions are called into action either because there are no formally or lexically transferable entities between the languages in question or because there are manifest constraints imposed by sentential or textual predilections pertaining to individual languages (Yip 1993). For example, in English, time concepts are often incorporated in the tenses of the verbs whereas in a non-morphological language like Chinese these time notions can only be rendered by lexical means. Similar peculiarities in either language are legion.

The following example shows how the translation has adopted the pre-modifiers *the-gods-mock-or-mar* and *not-know-themselves* by repeating the word *fool* and introducing the word *people* to cope with the original's post-modifiers, as Chinese syntax does not favour post-modifications of this kind:

- (3) The real fool, such as the gods mock or mar, is he who does not know himself<sup>[3f]</sup>.  
真正的傻瓜, 诸神用来取乐或取笑的傻瓜, 是那些没有自知之明的人<sup>[3f]</sup>.  
(lit.: *real fools, the-gods-mock-or-mar fools, are those not-know-themselves people*)

A large number of methodological strategies can be grouped under this category: *verb for e.g., abstract noun* in English; *clauses for phrases; active for passive; compound or complex sentences for simple sentences* in English; *time adverb or aspect particle* in Chinese for *tense forms* in English; *subjunctives* in English. Also, *substitution; synonymy; abbreviation; repetition; reiteration* to re-introduce the idea instead of repeating the word(s); *positive terms for negative ones or vice versa; clarity. Binomial or binominal; quadrisyllabic idiom* in Chinese; *pattern; rhythm; vernacular idiom* in Chinese; *idiomaticity; coinage; phonaetheme; reduplication* in Chinese; *domestication; foreignization; marked sentences* whose structure deviates from its usual arrangement; *unmarked sentences; relative clause* in English; *sentence adverb*; context-prompted *subject omission* in Chinese; and *imitation*, where the target language is seen to copy features from the source language.

#### 2.3.2.4. Information representation

If information realization in our proposition is more relevant to the peculiarities of a language's surface structure, then information representation is more interested in observing the choice of language-neutral, but discourse specific, means such as explanatory or metaphoric ways of verbalizing a message, which are universally apropos in all languages alike. For example, in

- (4) [When first I was put into prison some people advised me to try and forget who I was.] It was ruinous advice<sup>[76a]</sup>.  
[刚进监狱时, 有些人劝我忘掉自己是谁。] 要听了这话就完了<sup>[76a]</sup>.  
(lit.: *if [I] had listened to this utterance [I] would have been finished*)

where, as annotated in the corpus, the source text has been translated explanatorily.

The key methodological strategies include: *explanatory translation; metaphor for literal, metaphor, metonymy, personification*, attention to *connotation*, structural *variation, tone, and register*.

#### 2.3.2.5. Information explicitation

Arguably in translation there is the common tendency to make the source text information more explicit in the target text (Baker 1996), as can be testified to in the following example:

- (5) ... in the lowest mire of Malebolge I sit between Gilles de Retz and the Marquis de Sade<sup>[13c]</sup>.  
... 在地狱最底层的污浊中, 我将与崇拜撒旦的雷斯和性变态的萨德侯爵为伍<sup>[13c]</sup>.



(lit.: in the mire of the lowest level of hell, I will be with worship-Satan Gilles de Retz and sex-pervert Marquis de Sade)

It must nevertheless be noted that such explicitation as by adding modifiers may at times inadvisably highlight the headword and turn it into an unnecessary distraction at the text level. In this sense, *Gilles de Retz* and *the Marquis de Sade* have received more attention in the target text than in the source text. Another entity that may need explicitation in this case is *Malebolge*, which, in Dante Alighieri's *Inferno*, means the eighth circle of hell. But, as annotated, if the allusion had been spelt out in greater detail than it was, it would have diverted undue focus to itself and would have thus rendered the other explicitations (where the foci are supposed to be directed) less viable or effective.

Explicitation strategies identified in the corpus include: *attributive annotation*; *intratextual annotation*; *background knowledge*; *contextual awareness* based on *fore-going* or *following cotext*; *information retrieval*; *allusion*; and *etymology*.

### 2.3.2.6. Information and para-information

By *para-information* we mean the information that accompanies the material (i.e., connotative as well as denotative) content of a message – the information that is furnished by the distinctive form of the language itself. Para-information invariably induces a proactive tapping of the formal resources of a language for the desired textual effect. In translation, such tapping is by definition prompted by what is linguistically noticeable in the source text. For example, in a sentence like the following:

- (6) Pain, unlike Pleasure, wears no mask<sup>[87b]</sup>.

痛苦, 不象痛快, 是不戴面具的<sup>[87b]</sup>。

(lit.: pain, unlike joy, shi not wear mask de)

where, in terms of information organization, the emphasized rheme *wears no mask* (compare: *does not wear a mask*) is similarly underscored in the Chinese text by a marked *shi...de* structure (compare: 没戴面具 literally *have-not wear mask*). What is of interest in this connection, however, is the para-information (i.e., contrast) carried in the alliterative *pain* and *pleasure*, which has prodded the translator to tap the Chinese language for potential devices to register a similar contrast. Hence 痛苦 (*pain* pronounced *tongku*) and 痛快 (*joy* pronounced *tongkuai*) with an identical first syllable *tong* to evoke a matching alliterative effect. A version perhaps more true to the material content of *Pleasure* 享受 (*pleasure* pronounced *xiangshou*) would have otherwise mitigated the effect.

Methodological strategies to recover such accompanying para-information are annotated under such searchwords as: *prosody*; *rhythm*; *symmetry*; *parallelism*; *contrast*; *progression*; *pun*; *alliteration*; *rhyme*; *cadence*; and *proverb*.

### 2.3.2.7. Information reformulation

Information reformulation focuses on the alternative perspectives from which the target text features a given message in its information structuring. That is to say, a translator may choose to base the translation on visualized images of the characters or incidents in the story rather than on the actual wording that the story-teller uses. Such reformulation, by virtue of the details of images and events captured, will result



in a reintegration of form and content in its textualization. A whole spectrum of external or internal forces, social, political, economical, moral, legal, religious, ethnic, aesthetic, and so on, may influence the translator's choice of language and hence his methodological decisions in reformulating the information flow. For example, we can see in the following passage the translation, instead of adhering verbally to the original wording, chooses to register a series of images in a story-telling sequence that appeals to readers in the Chinese language:

- (7) [...] or that evil passions fled at his approach, and men whose dull unimaginative lives had been but a mode of death rose as it were from the grave when he called them: or that when he taught on the hillside the multitude forgot their hunger and thirst and the cares of this world, and that to his friends who listened to him as he sat at meal the coarse food seemed delicate, and the water had the taste of good wine, and the whole house became full of the odour and sweetness of nard<sup>[98b]</sup>.

[...] 走来, 罪恶的情欲便遁逃无影, 而那些生活暗淡想象力阙如的人们, 便象死而复生一样, 从坟墓中随他的召唤站立起来; 他在山边讲道, 众人听着便忘了饥渴和尘世的纷扰; 他坐在餐桌边, 聆听他教诲的朋友们便觉得粗茶淡饭也成了美味佳肴, 白水喝着尤如美酒, 整个屋子弥漫着甘松油的甜香<sup>[98b]</sup>。

(lit.: he came, evil passions fled without a trace, and those dull-living unimaginative people, as if revived from death, from their graves following his call rose; he on the hillside preached, the multitude listened and forgot their hunger and thirst and their worldly cares; he sat at the meal table, listening-to-his-teaching friends found the coarse food to be delectable delicacies, plain water tasted like beautiful wine, the whole room was permeated with the sweet fragrance of nard)

Strategies of reformulation include: *narrative translation* that may override the actual event-sequence in the source text; *descriptive translation*, where images rather than words dominate; *intensity* of wording; and *flexibility*.

### 2.3.3. A glossary for users

A glossary derived from our contemplation above is included in the system to give the user a schematic picture of the searchwords used in the corpus. While the categorization suggests that the translation methodologies were not identified and annotated at random, a note of caution seems in order here: the categorization is meant to serve as a useful template for reference rather than a clear-cut demarcation for prescriptive determination. Also, the exemplification in this report is illustrative but not exhaustive, for to map out all the methodological or phenomenological complexities of translation is virtually a mission impossible. Translation, in Richards's (1953: 250) much quoted words, "may very probably be the most complex type of event yet produced in the evolution of the cosmos."

Indeed, as our data confirms, language in use can be clearly seen as an individual (or idiosyncratic) way of encoding a message, which, in most cases and conspicuously so in literary contexts, imparts not only information in terms of specific knowledge, cultural underpinnings, the encoder's political stance and his social, geographical or perhaps religious affiliation but also what we call para-information in terms of the encoder's favoured linguistic style or sometimes conscious artistic manipulations of language itself. Consequently, the list of searchwords generated from this project remains open and is to be extended as an overarching inventory to accommodate the translation of other types of text and by other translators.

### 3. The use of the system as an auxiliary teaching/learning tool

As a pilot project, the software has been distributed to students on under- and post-graduate courses taught by the first author, such as translation methodology and stylistics and translation as a supplement to class materials. It has enabled the teacher to assign tasks requiring the students to consult the corpus and prepare their discussion before coming to class. In our trial runs of the system, the students were given a group of related searchwords to prepare before the class, where the lecturer could initiate more in-depth demonstrations and discussions. Such class discussions could be followed by similarly searchword-led tutorials. Discussions with a common focus have produced in the tutorials a more active and better prepared involvement on the part of the students. More significantly, such a software programme in the long run will help students wean themselves off being spoon-fed and exercise control over what they need to learn, when and at what pace, the potential of which, as we can see, will be more readily appreciated in distance-learning, self-study, and web-teaching contexts. Further to the above brief account, more extensive tests on the system's efficiency can be conducted, and more systematic feedback expected with its publication in China in book plus CD-ROM form (Zhu 2008).<sup>4</sup>

Access to these self-learning devices does not of course completely replace face-to-face tuition or hands-on practice in translation teaching. What it does offer, however, is greater flexibility in learning, a research-informed consistency of explaining translation methodologies and phenomena, and the possibility of a self-monitored progression.

### 4. Concluding remarks

A translator's experience might be unique (or even idiosyncratic) in itself. However, once the underlying patterns beneath the seemingly idiosyncratic surface are systematically identified, characterized in a theoretically-accountable manner, and then clearly explained in relation to their applicability in actual translation, they become useful resources of translation pedagogy, which are universally transferable and can be readily tapped and re-tapped to enrich the experience of students of translation or newcomers to the profession. And the descriptive nature of this mine of methodological information will be a veritable antidote to impressionistic, *ad hoc*, and prescriptive views or approaches, or *naïve empiricism* in Neubert and Shreve's words (1992: 33; see also Zhu 2002) based on "entirely subjective descriptions of 'what has worked for [the pedagogues].'" That is the pedagogical goal the project has been trying to achieve.

As a way of teaching translation, the project is innovative in the sense that it provides a theory-informed characterization of translation strategies for description, discussion, and explanation of actual translation acts. The findings are then built into a user-friendly pedagogical software programme, complete with instructions for use and a list of searchwords indicating methodological strategies in translation, readily consultable on a computer CD-ROM with a view to helping students to learn how to learn, when dealing with materials of their interest.

All in all, this project can be seen as representing an innovative effort at applying theory and corpus-based approaches to the practice of translation and translator

“training in the three different but interrelated directions” as identified by Lederer (2007: 29), namely,

- a) introducing students to translation methodology;
- b) rationalizing their didactic progression through a judicious selection of teaching materials;
- c) assessing trainees’ work.

Looking ahead, we are contemplating further development of the system on the following two fronts. First, it can be made more user-friendly in design, more self-sufficient by including components such as guided exercises and translation comparisons, as well as facilities for learner-tutor communication. Secondly, the planned series stemming from this pilot project will as a whole feature a greater variety of text types, subject matter, translators, and language combinations with more refined annotations of methodologies and related phenomena of an interlingual nature.

#### ACKNOWLEDGEMENTS

The project described in this report has been fully supported by two grants from the City University of Hong Kong (Project Nos. 7100174 and 7001724). The first author also wishes to thank La Trobe University, Australia, for an IAS Fellowship (June–August 2006), which has helped expedite the completion of this report. The project was first presented at the international symposium, *For a Proactive Translationology*, Université de Montréal, 7–9 April 2005.

#### NOTES

1. The news was published in *Chinese Translators Journal* 27(3) (2006), p. 85. Figures from other sources show that in 2005 there were 3,975 out of the 17,704 candidates obtaining the professional translation certificates. Also, it was reported that in mainland China there were more than 3,000 registered translation agencies with revenues of US\$2.5 billion in 2005, and that translation in China, an industry employing more than 60,000 full-time translators with “several more hundreds of thousands” working part-time or as freelancers, is in need of nation-wide rigorous assessment mechanisms to “supervise the quality of translation work and regulate the market,” see Xiu’s (2006) keynote speech at the First China International Forum on Translation Industry, Shanghai, 2006, also the English news reports, one by Yan (2006) and one published in *People’s Daily Online* (Unknown author 2006).
2. The text was originally solicited by People’s Literature Publishing House in Beijing in 1997 for inclusion in its *Works of Oscar Wilde* edited by Fuzhong Su (Zhu 2000). Another prestigious publisher in China, Chinese Literature, later bought the translation from the first publisher to replace the one it had already commissioned for its *Complete Works of Oscar Wilde* (Zhu 2001; personal letter from Su 6 November 1998).
3. Examples used in this article to illustrate various points are all from Zhu (2008), retaining the square-bracketed markers in superscript which indicate their respective position in the bilingual corpus (see Sub-Section 2.1.2 and Table 1 in Appendix for further explanation). The “literal” back translations are provided by the present authors for reference.
4. At the Summer Symposium on Translation and Interpreting organized by the Translators Association of China (Beijing, 19–25 July 2007), we conducted a preliminary questionnaire survey following an extended presentation on the conceptual, structural, and operational features of the software programme. Fifteen symposium participants responded to the survey. The respondents were from various universities across the country, with professional background ranging from fresh graduate (1), teacher of English (8), to teacher of translation (6), eight of them currently working on an MA programme and four having an interest in translation. Specific statistics are reported as follows (the figures in brackets indicate the numbers of the respondents who have ticked the choices off):
  - a) The selection of the data is (only one of the 4 choices can be ticked off): very inappropriate (0), inappropriate (0), appropriate (11), very appropriate (4);

- b) The annotations are (more than one of the 11 choices can be ticked off): too simple (0), simple (0), appropriate (8), complicated (1), too complicated (2), not convincing (0), convincing (8), very convincing (2), inconsistent (1), consistent (7), very consistent (1);
- c) The retrieval methods are (only one of the 5 choices can be ticked off): very difficult to learn (0), difficult to learn (0), suitable (9), easy to use (4), very easy to use (2);
- d) The software programme is suitable for (more than one of the 9 choices can be ticked off): senior middle school English classes (0), translation classes of non-English majors (4), translation classes of English majors (9), BA programmes in translation (12), postgraduate programmes in translation (9), for research purposes as a corpus (8), long-distance teaching of translation (10), self-study (9), out-of-class bilingual reading (4), leisure reading in literature (5);
- e) I will (more than one of the 14 choices can be ticked off): not use it (0), use it (4), use it regularly (3), concentrate on using it for a period of time (5), not recommend it to friends (0), recommend it to friends (11), not recommend it to undergraduates (0), recommend it to undergraduates (8), not recommend it to postgraduate (0), recommend it to postgraduates (5), not use it in the classroom (0), use it in the classroom (6), not use it as corpus for research (0), use it as corpus for research (9). The respondents' free comments on features they like most identified the following: authenticity of data, completeness of system (4), richness of data (2), convenience (7), annotations of background knowledge (2), flexibility (1), systematicity (1), dual annotation (1), searchword retrieval (1), and thought-provokingness (1).

The respondents' suggestions for improvement were related to the following aspects: incorporation of theories beyond information structuring in its explanatory framework, inclusion of more text pairs, attention to textualization at the discourse level, enhanced accuracy in annotation and the use of data, conciseness of annotation, enhanced teacher-friendliness, clearer reference to the theoretical concepts used, the issue of open-endedness of literary translation, and the possibility of on-line access.

## REFERENCES

- BAKER, Mona (1993): Corpus linguistics and translation studies – implications and applications. In: Mona BAKER, Gill FRANCIS and Elena TOGNINI-BONELLI, eds. *Text and Technology – In Honour of John Sinclair*. Amsterdam: John Benjamins, 233-250.
- BAKER, Mona (1996): Corpus-based translation studies: the challenges that lie ahead. In: Harold SOMERS, ed. *Terminology, LSP and Translation: Studies in Language Engineering – In Honour of Juan C. Sager*. Amsterdam: John Benjamins, 175-186.
- BOWKER, Lynne (2001): Towards a methodology for a corpus-based approach to translation evaluation. *Meta*. 46(2):345-364.
- BRISSET, Annie (2003): Alterity in translation: an overview of theories and practices. In: Susan PETRILLI, ed. *Translation Translation*. Amsterdam: Rodopi, 102-132.
- CARTER, Ronald and NASH, Walter (1983): Language and literariness. *Prose Studies*. 6(2):123-141.
- COOK, Guy (1994/1999): *Discourse and Literature: The Interplay of Form and Mind*. Oxford/Shanghai: Oxford University Press/Shanghai Foreign Language Education Press.
- HALLIDAY, Michael A. K. (1994): *An Introduction to Functional Grammar*. London: Arnold.
- HART-DAVIS, Rupert (1962): *The Letters of Oscar Wilde*. London: Hart-Davis.
- HYDE, H. Montgomery (1982): *The Annotated Oscar Wilde*. London: Orbis.
- KENNY, Dorothy (1998): Corpora in translation studies. In: Mona BAKER, ed. *Routledge Encyclopedia of Translation Studies*. London: Routledge, 50-53.
- KENNY, Dorothy (2001): *Lexis and Creativity in Translation: A Corpus-based Study*. Manchester: St. Jerome.
- KIRALY, Don (2000): *A Social Constructivist Approach to Translator Education; Empowerment from Theory to Practice*. Manchester: St. Jerome.
- LAMBRECHT, Knud (1994): *Information Structure and Sentence Form: Topic, Focus, and the Mental Representation of Discourse Referents*. Cambridge: Cambridge University Press.
- LAVIOSA, Sara (1998): The corpus-based approach: a new paradigm in translation studies. *Meta*. 43(4):474-479.

- LEDERER, Marianne (2007): Can theory help translator and interpreter trainers and trainees? *The Interpreter and Translator Trainer*. 1(1):15-35.
- LEECH, Geoffrey N. and SHORT, Michael H. (1981): *Style in Fiction*. London: Longman.
- LITTLE, David (2000): Autonomy and autonomous learners. In: Michael BYRAM, ed. *Routledge Encyclopedia of Language Teaching and Learning*. London: Routledge, 69-72.
- LIU, Xiliang (2006): Ensuring sustainable development of China's translation industry through regulatory measures: Possible solutions to the problems currently besetting China's translation industry. *Chinese Translators Journal*. 27(4):5-7.
- MCCARTHY, Michael (1991): *Discourse Analysis for Language Teachers*. Cambridge: Cambridge University Press.
- NEUBERT, Albrecht and SHREVE, Gregory M. (1992): *Translation as Text*. Kent: Kent State University Press.
- RICHARDS, Ivor A. (1953): Toward a theory of translating. In: Arthur F. WRIGHT, ed. *Studies in Chinese Thought*. Chicago: University of Chicago Press, 247-262.
- Unknown author (2006): China's 60,000 professional translators cannot cope with market demand. *People's Daily Online*. 30 May 2006. Visited 3 June 2006, <<http://english.people.com.cn/>>.
- WILDE, Oscar (1897/1951): *De Profundis: Being the first complete and accurate version of 'Epistola: in carcere et vinculis' the last prose work in English*. Vyvyan HOLLAND, ed. London: Methuen.
- YAN, Zhen (2006): Translation to be supervised. *Shanghai Daily*, 29 May 2006. Visited 3 June 2006, <[http://www.shanghaidaily.com/art/2006/05/29/280325/Translation to be supervised.htm](http://www.shanghaidaily.com/art/2006/05/29/280325/Translation%20to%20be%20supervised.htm)>.
- YIP, Po-Ching (1993): Linguistic dispositions – a macroscopic view of translating from English into Chinese. *Journal of the Chinese Language Teachers Association*. 28(2):25-74.
- ZHU, Chunshen (1996a): Syntactic status of the agent and information presentation in translating the passive between Chinese and English. *Multilingua*. 15(4):397-417.
- ZHU, Chunshen (1996b): Translation of modifications: about information, intention and effect. *Target*. 8(2):301-324.
- ZHU, Chunshen (1999): UT once more: the sentence as the key functional Unit of Translation. *Meta*. 44(3):429-447.
- ZHU, Chunshen (2000): 自深深处 In: Fuzhong SU, ed. 王尔德作品集 (*Works of Oscar Wilde*). Beijing: People's Literature.
- ZHU, Chunshen (2001): 致阿尔弗雷德·道格拉斯勋爵. In: Zhao WEN, Shengkang Ji *et al.* eds. 王尔德全集6, 书信(下) (*Complete Works of Oscar Wilde, Vol. 6, letters II*). Beijing: Chinese Literature, 59-185.
- ZHU, Chunshen (2002): Translation: theories, practice, and teaching. In: Eva HUNG, ed. *Teaching Translation and Interpreting 4*. Amsterdam: John Benjamins, 19-30.
- ZHU, Chunshen (2004): Repetition and signification: a study of textual accountability and perlocutionary effect in literary translation. *Target*. 16(2):227-252.
- ZHU, Chunshen (2005): Machine-aided teaching of English-Chinese financial translation. Education development project for internal teaching purposes. City University of Hong Kong, Project No. 6980030.
- ZHU, Chunshen (2008): 自深深处 (Bilingual electronic edition with the original text *De Profundis* by Oscar Wilde, including machine retrievable annotations on background knowledge and translation techniques, and soundtracks in Chinese and English). Nanjing: Yilin Press.

APPENDIX

TABLE 1

Illustration of the dual mode of annotation

<p>[23] On your return to town from the actual scene of the tragedy to which you had been summoned, you came at once to me very sweetly and very simply, in your suit of woe, and with your eyes dim with tears. You sought consolation and help, as a child might seek it. I opened to you my house, my home, my heart. <b>I made your sorrow mine also, that you might have help in bearing it</b>[23a]. <b>Never, even by one word, did I allude to your conduct towards me, to the revolting scenes, and the revolting letter</b>[23b]. Your grief, which was real, seemed to me to bring you nearer to me than you had ever been. The flowers you took from me to put on your brother's grave were to be a symbol not merely of the beauty of his life, but of the beauty that in all lives lies dormant and may be brought to light.</p> <p>[24] <b>The gods are strange. It is not of our vices only they make instruments to scourge us.</b>[24.1] <b>They bring us to ruin through what in us is good, gentle, humane, loving</b>[24a]. <b>But for my pity and affection for you and yours, I would not now be weeping in this terrible place</b>[24b].</p>	<p>[23] 从他们传召你去的悲剧现场一回到城里,你马上就到我这儿来,穿着丧服,泪眼盈盈的一派温良率真的模样,要人安慰、求人帮忙,象小孩一样。我对你敞开了我的房子,我的家,我的心。将你的悲痛当作自己的悲痛,这样也许能在你的沉沉哀痛中扶你一把[23a]。我甚至绝口不提你是怎么待我的,绝口不提那一幕幕不堪入目的吵闹和那一封不堪入耳的信[23b]。你那真切的悲哀,似乎带着你前所未有的靠近我。你从我这儿带去供在你哥哥坟上的鲜花,不止要成为他生命之美的象征,也要成为蕴藏于所有生命中并可能绽放的美的象征。</p> <p>[24] 神是奇怪的。他们不但借助我们的恶来惩罚我们,也利用我们内心的美好、善良、慈悲、关爱,来毁灭我们[24a]。要不是因为我对你及你家人的怜悯和感情,我现在也不会在这人所不齿的地方哭泣[24b]。</p>
--	--

FIGURE 1  
Retrieving annotation from the Searchword List

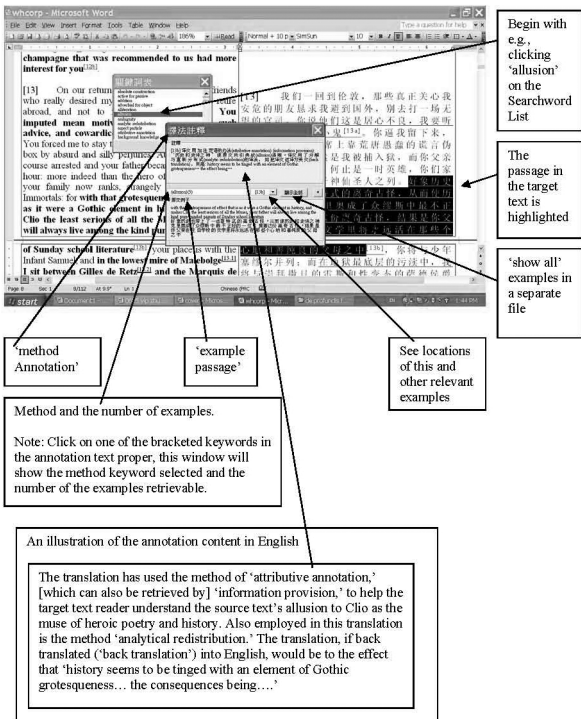




FIGURE 2  
Retrieving background knowledge annotation

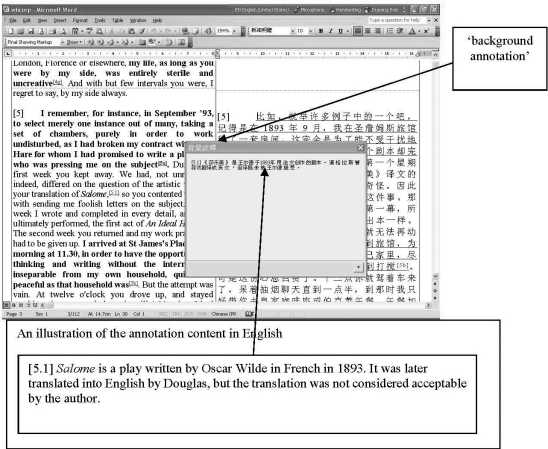


FIGURE 3  
Example of retrieving annotations from the corpus

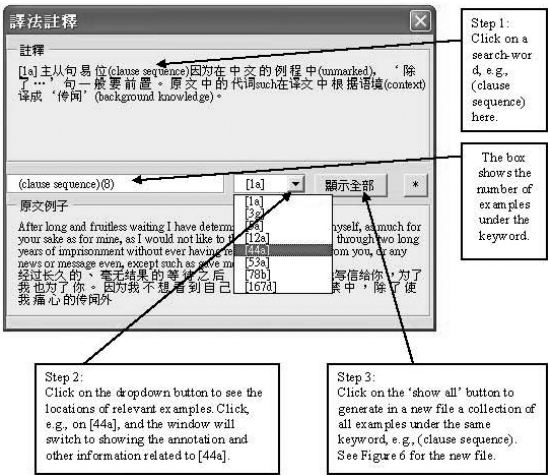




FIGURE 4  
Clicking back from the ‘show all’ new file to the corpus proper

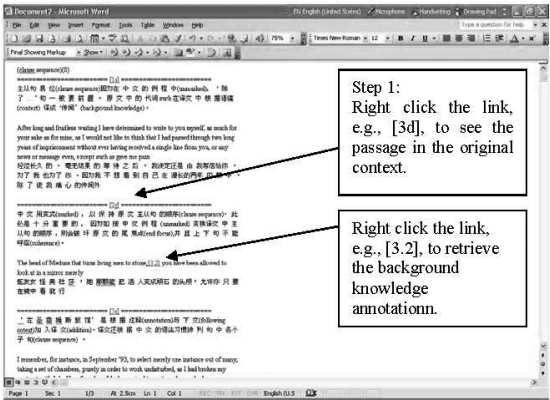


FIGURE 5  
The screen after clicking back from the ‘show all’ file

